

CUSTOMER DATA ANALYSIS — FINAL REPORT

1. Title Page

CUSTOMER DATA ANALYSIS

Project ID: PTID-CDA-NOV-25-880

Project Code: PRDA-05

Prepared by: Ankit Jha (14TH July CDA Batch)

2. Introduction

This project focuses on identifying customer purchasing patterns using a combined approach of Python-based Exploratory Data Analysis (EDA) and a Power BI dashboard.

The goal is to understand **who the customers are, what they buy, how they prefer to pay, and which demographic contributes most to revenue.**

The project helps uncover actionable business insights for marketing, inventory planning, store targeting, and customer segmentation.

3. Project Objectives

1. Analyze customer demographics (gender, age group).
 2. Examine shopping behavior across categories and malls.
 3. Understand payment method preferences.
 4. Evaluate revenue contribution by customer segments.
 5. Build an interactive Power BI dashboard for business insights.
 6. Provide data-driven recommendations to the company.
-

4. Dataset Description

The dataset contains customer transaction details including:

- Customer ID
- Gender
- Age
- Category of purchase
- Quantity purchased
- Shopping mall
- Payment method

Derived columns created during analysis:

- **Revenue** = quantity × price
 - **Age Group** (0–18, 19–25, 26–35, 36–45, 46–60, 60+)
-

5. Methodology

5.1 Data Cleaning (Python)

- Removed duplicates
- Checked and handled missing values
- Converted data types
- Created new fields (revenue, age_group)
- Corrected inconsistencies in categories

5.2 Exploratory Data Analysis

Using Python (pandas, matplotlib):

- GroupBy analysis for revenue and quantity
- Crosstab analysis for segment relationships
- Visualization of demographic patterns
- Category-wise performance
- Payment method comparisons

5.3 Dashboard Development (Power BI)

- Loaded cleaned dataset (customer_cleaned.csv)
 - Added slicers for interactive filtering
 - Designed charts for gender, age, category, mall, and payment trends
 - Organized layout into professional sections
 - Applied clean theme and color palette
 - Finalized one-page interactive dashboard
-

6. Python (Jupyter) Analysis Summary

6.1 Gender Analysis

- Female customers purchased more products (178,659 items).
- Female customers generated more revenue (₹150M).
- Female customers dominate across most categories.

6.2 Age Group Analysis

- Age group **46–60** purchased the most items and generated the highest revenue.
- Age group **60+** also has high revenue contribution.
- Young age group (0–18) is the least active.

6.3 Category Insights

- Clothing has the highest total item sales.
- Technology generates high average revenue per purchase.
- Cosmetics, Food & Beverage, and Shoes perform strongly in volume.

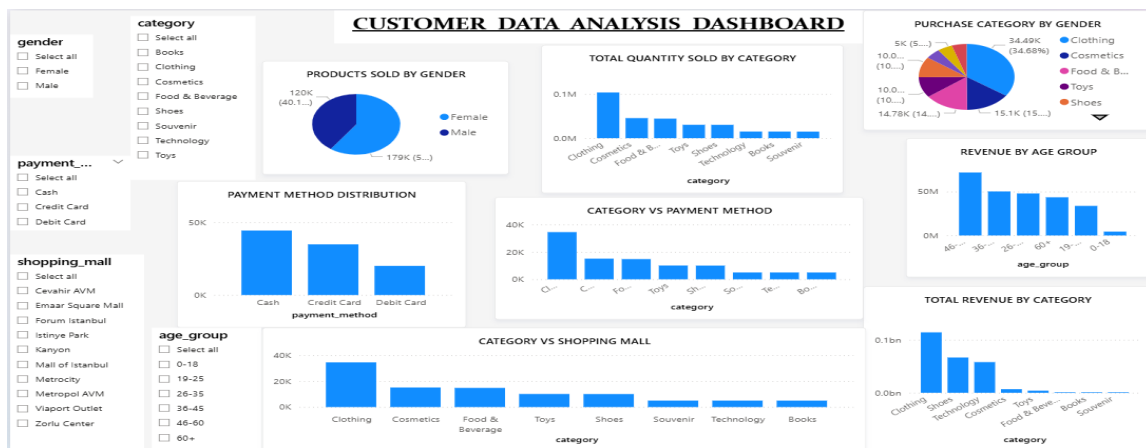
6.4 Payment Method Analysis

- Cash is the most frequently used method.
- Credit cards dominate in high-value categories.
- Debit cards remain steady but least used.

6.5 Shopping Mall Insights

- Mall of Istanbul and Kanyon show high transaction activity.
- Premium malls have higher credit-card usage and bigger purchases.

7. Power BI Dashboard Overview



The dashboard includes:

Key Visuals:

- Payment Method Distribution
- Products Sold by Gender
- Revenue by Age Group
- Quantity Sold by Category

- Category-wise Revenue
- Category vs Gender
- Category vs Payment Method
- Category vs Shopping Mall

Filters (Slicers):

- Gender
- Age Group
- Category
- Payment Method
- Shopping Mall

The dashboard allows users to explore trends intuitively and supports decision-making through visual storytelling.

8. Insights Summary

1. **Female customers are the strongest customer segment** in both quantity purchased and revenue generated.
 2. **Age group 46–60 contributes the highest revenue**, making them the primary consumer base.
 3. **Clothing is the top-selling category**, while **Technology** has high-value purchases.
 4. **Cash remains the most used payment method**, but credit cards drive premium purchases.
 5. **Premium malls generate more revenue**, especially from card payments.
 6. **Food & Beverage and Cosmetics** show strong volume, indicating impulse-driven buying.
 7. Young customers (0–18) are the least contributing segment.
-

9. Recommendations to the Company

1. **Target female customers** with loyalty programs and personalized promotions.
2. **Focus marketing on ages 36–60**, especially with technology and apparel categories.
3. **Increase credit card partnership offers**, as card users tend to make higher-value purchases.
4. **Strengthen premium mall presence**, as they generate stronger revenue.
5. **Bundle high-performing categories** like Clothing, Shoes, and Cosmetics for festive promotions.
6. **Optimize inventory** in top-performing malls and age groups.

7. Introduce **age-specific and mall-specific offers** to increase conversions.
-

10. Executive Summary

This project analyzes customer shopping behavior across various malls using Python and Power BI. The analysis reveals clear differences in how gender, age groups, and shopping categories influence revenue. Female customers and the age group 46–60 form the highest revenue-generating segment. Clothing is the most purchased category, while technology has the highest value per purchase.

Payment behavior shows that cash is still dominant, but credit card usage is strongly linked with high-value transactions. Premium malls contribute significantly to revenue through card-based purchases. The final Power BI dashboard visually summarizes these findings, enabling leadership to take strategic decisions related to marketing, inventory, and customer engagement.
